**Programming with R**

**Assignment – 1**

**Max. Marks: 40 Due Date: 29 Jan, 2020**

**Note: Assignment to be submitted in Blue Book only**

1. Write a note on how dplyr package helps in data transformation. Explain briefly the below key dplyr functions that help in data manipulation.

**(18)**

filter()

arrange()

select() and rename()

mutate() and transmute()

summarise()

sample\_n() and sample\_frac()

Resource – Refer the Introduction to dplyr Vignette available on cran

(<https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html>)

1. Download the Placement\_Data\_Class.csv file from Github and answer the following questions. For every question write the code followed by the answer. Use the dplyr functions only. (<https://github.com/DG1606/CMS-R-2020>)

**(22)**

Data Description:

ssc\_p – Percentage in Class 10

ssc\_b – Board studied in Class 10

hsc\_p – Percentage in Class 12

hsc\_b – Board studied in Class 12

hsc\_s – Stream studied in Class 12

degree\_p – Percentage in Degree

degree\_t – Type of Degree (Science & Technology / Commerce & Management / Others)

workex – Work experience

etest\_p – Percentage in Entrance Test

status – Placement status

1. How will you filter students from Science and Technology (Sci&Tech) degree (degree\_t) having degree percentage (degree\_p) ≥ 75. How many such students are there?
2. From the above filter students with gender as F. How many obse5rvations do you get?
3. How will you select a subset of the data of students with degree\_p ≤ 52? Assign the dataset to p\_lower. How many observations does p\_lower contain? Use this dataset to answer the remaining questions.
4. Reorder (ascending) the rows in p\_lower based on etest\_p. What is the sl.no of the student at the top and bottom?
5. Reorder (descending) the rows in p\_lower based on ssc\_p. What is the sl.no of the student at the top and bottom?
6. How will you select columns that ends with \_p? How many columns got selected?
7. How will you rename the degree\_p column as degreepercent?
8. Create a new variable (avg\_p) which shows the average percentage of ssc\_p and hsc\_p. What is the avg\_p of the first student in the output?
9. How will you display only the new variable (avg\_p) in the output in case of above?
10. How will you calculate/summarise the median of degree\_p of all students? What is the median value?
11. How will you select a random sample of 2 students? Use seed = 25. What are the sl.no of the selected students?